

CS 134: Networks

Problem Set 6

Xiner Zhou

March 20, 2017

1. (30 points)

- a. **Solution:** The main difference between the variant of the branching process and the one we discussed in lecture is that, in lecture we assumed nodes get activated independently and identically as a series of Bernoulli events, but now we have no such independent Bernoulli probabilistic assumption anymore, instead, we only assume that for any active nodes v at level n , the number of its active children is a random variable $Z_v \sim P(l)$, i.i.d., and all sets of active children that have the same size are equally likely.

For $n \geq 1$, the number of active nodes at level n is equal to the total number of active children of all nodes at level $n-1$, and there are k^{n-1} nodes at level $n-1$. Therefore, we have:

$$\begin{aligned}
 E[X_n] &= E[\text{number of active nodes at level } n] \\
 &= E\left[\sum_{j=1}^{k^{n-1}} \text{active children of } v_j \text{ at level } n-1\right] \\
 &= E\left[\sum_{j=1}^{k^{n-1}} Z_{v_j}\right] \\
 &= \sum_{j=1}^{k^{n-1}} E[Z_{v_j}]
 \end{aligned}$$

Since all Z_{v_j} are independently, identically distributed as $P(l)$, where $l \in 1, \dots, k$, by the definition of expectation of a discrete random variable:

$$\begin{aligned}
 &= \sum_{j=1}^{k^{n-1}} [0 \times P(0) + 1 \times P(1) + \dots + k \times P(k)] \\
 &= k^{n-1} \sum_{i=1}^k [iP(i)]
 \end{aligned}$$

- b. **Solution:** Let $q_n = Pr[X_n \neq 0]$, since the root node r (the sole node at level $n = 0$) is assumed to be active, therefore, $q_0 = 1$.

c. **Solution:** q_n is decreasing in n . We can break down the event that there are at least 1 active node at phase n into two stages. In order to have the branching process still persists at level n , it would require : (1) There are at least 1 active node at phase $n-1$, otherwise, the whole branching process would stop; and (2) There are at least 1 active node at phase $n-1$ have at least 1 active children. The probability of the second stage is always less than 1, and since $q_n = \Pr[\text{first stage}] \times \Pr[\text{second stage}] = q_{n-1} \times \Pr[\text{second stage}]$, it is clear that $q_n < q_{n-1}$.

d. **Solution:** Consider a node j in phase 1 of the contagion model with parameter k and a probabilistic distribution $P(i)$. The probability that node j is active is:

$$\begin{aligned}\Pr[\text{node } j \text{ is active}] &= \sum_{i=1}^k P(i) \times \frac{\binom{k-1}{i-1}}{\binom{k}{i}} \\ &= \sum_{i=1}^k P(i) \frac{i}{k} \\ &= \frac{\sum_{i=1}^k i P(i)}{k}\end{aligned}$$

The probability that there exists at least one active node at phase n that is activated from a branch through j is then $\frac{\sum_{i=1}^k i P(i)}{k} q_{n-1}$.

Then the probability that no node is active at phase n that is activated from a branch through j is $1 - \frac{\sum_{i=1}^k i P(i)}{k} q_{n-1}$.

Given that there are k such nodes j at phase 1, the total probability that there is no active node at phase n is $(1 - \frac{\sum_{i=1}^k i P(i)}{k} q_{n-1})^k$.

Thus, the probability that there exists at least one active node at phase n is,

$$q_n = 1 - (1 - \frac{\sum_{i=1}^k i P(i)}{k} q_{n-1})^k$$

We now have a recursive formula for $q_n = f(q_{n-1})$, which is,

$$f(x) = 1 - (1 - \frac{\sum_{i=1}^k i P(i)}{k} x)^k$$

Notice that $\sum_{i=1}^k i P(i) = E[Z_v]$ is the expected number of active children for any active nodes, we can also write as $f(x) = 1 - (1 - \frac{E[Z_v]}{k} x)^k$.

e. **Solution:** The first derivative:

$$\begin{aligned}f'(x) &= k \times \frac{\sum_{i=1}^k i P(i)}{k} \times (1 - \frac{\sum_{i=1}^k i P(i)}{k} x)^{k-1} \\ &= \sum_{i=1}^k i P(i) (1 - \frac{\sum_{i=1}^k i P(i)}{k} x)^{k-1}\end{aligned}$$

Since $\sum_{i=1}^k i P(i) = E[Z_v]$,

$$= E[Z_v] (1 - \frac{E[Z_v]}{k} x)^{k-1}$$

Since $Z_v \in 1, \dots, k$ and $0 < P(0) < 1$, $E[Z_v] < k$. Therefore, $\frac{E[Z_v]}{k} < 1$.

Then, since $\frac{E[Z_v]}{k} < 1$ and the domain $x = [0, 1]$, we have that the first derivative is positive

$$f'(x) > 0$$

The second derivative:

$$\begin{aligned} f''(x) &= \sum_{i=1}^k iP(i) \times (k-1) \times \left(-\frac{\sum_{i=1}^k iP(i)}{k}\right) \times \left(1 - \frac{\sum_{i=1}^k iP(i)}{k}x\right)^{k-2} \\ &= -(k-1)E[Z_v]\frac{E[Z_v]}{k}(1 - \frac{E[Z_v]}{k}x)^{k-2} \end{aligned}$$

For the same reason as above, we have that the second derivative is negative

$$f''(x) < 0$$

Taken together, we can conclude that $f'(x)$ is a positive and decreasing function, and that $f(x)$ is a concave down function from $x = 0$ to $x = 1$.

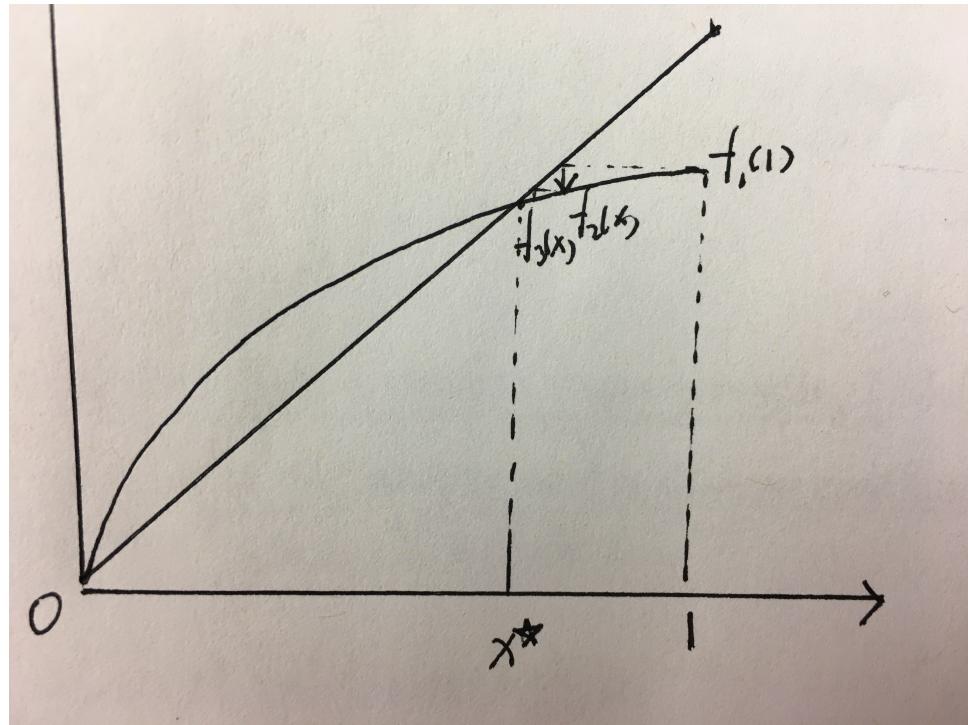
f. Solution:

To find $q_{\inf} = \lim_{n \rightarrow \infty} q_n$, we first look at some properties of $f(x)$:

- $f(0) = 0$, and $f(1) = 1 - (1 - \frac{\sum_{i=1}^k iP(i)}{k})^k \in (0, 1)$, since $\sum_{i=1}^k iP(i) = E[Z_v] \in (0, k)$
- $f(x)$ is a concave down function from $x = 0$ to $x = 1$, from part (e)
- $f'(0) = \sum_{i=1}^k iP(i) = E[Z_v]$

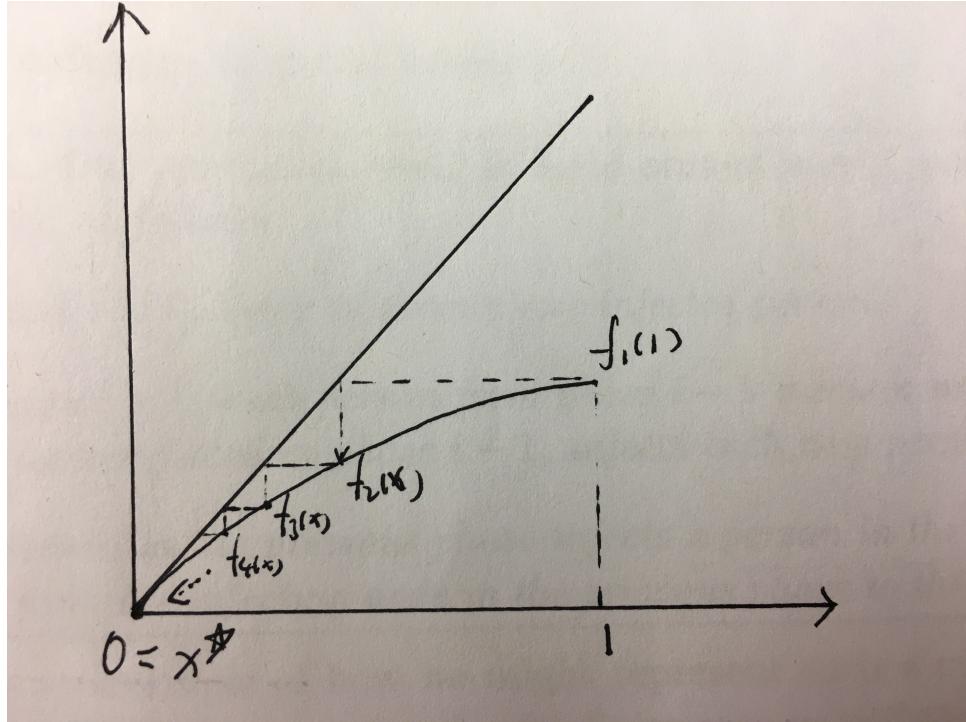
Given these three properties, we plot the function $f(x)$ and $y = x$, restricting domain to $x = [0, 1]$ since x represents a probability. The plot depends on whether $f'(0) = \sum_{i=1}^k iP(i) = E[Z_v]$ is greater than 1, or less than 1.

Scenario 1: If $f'(0) = \sum_{i=1}^k iP(i) = E[Z_v] > 1$. Since $f(0) = 0$ and $f'(0) > 1$, for some small positive values of x , $f(x)$ curve will be above $y = x$; however, since $f(1) < 1$, we know that at $x = 1$, $f(x)$ curve will be below $y = x$, producing the concave plot as below.



To see the limiting behavior, we start from $(1, f(1))$ and move horizontally to the $y = x$ curve so that we are at $(f(x), f(x))$. If we continue this process indefinitely, we will eventually stop when $y = x$ and $y = f(x)$ curves meet; this point of intersection, (x^*, x^*) is the convergence of the sequence q_0, q_1, q_2, \dots and thus we find that $q_{\inf} = \lim_{n \rightarrow \infty} q_n > 0$, as the point (x^*, x^*) is strictly between 0 and 1.

Scenario 2: If $f'(0) = \sum_{i=1}^k iP(i) = E[Z_v] < 1$. Since $f(0) = 0$, $f'(0) < 1$, and $f(1) < 1$, we know that $f(x)$ curve will always be below $y = x$, producing the concave plot as below.



We find that $q_{\inf} = \lim_{n \rightarrow \infty} q_n = 0$, as the intersection point (x^*, x^*) is at $(0, 0)$.

To summarize, when $\sum_{i=1}^k iP(i) > 1$, or the expected number of active children of active nodes $E[Z_v] > 1$, the disease will never die out with probability greater than 0; when $\sum_{i=1}^k iP(i) < 1$, or the expected number of active children of active nodes $E[Z_v] < 1$, the disease will eventually stop spreading with probability 1. Informally, it is saying that we should worry about a disease becoming epidemic if and only if a typical carrier infects more than one new healthy individual. This is consistent with the results in lecture.

- g. Solution:** I have given a full prove in (f) that, the necessary and sufficient condition for $q_{\inf} > 0$ is that, $\sum_{i=1}^k iP(i) > 1$, or the expected number of active children of active nodes $E[Z_v] > 1$. Informally, it is saying that we should worry about a disease becoming epidemic if and only if a typical carrier infects more than one new healthy individual.

2. (20 points)

- a. Solution:** If $r : 0, \dots, k \rightarrow \mathbb{R}$ is a strictly decreasing function, then $-r$ is a strictly increasing function. Since \widetilde{P} first-order stochastic dominates P , we have that

$$\begin{aligned} \sum_{l=0}^k \widetilde{P}(l)[-r(l)] &> \sum_{l=0}^k P(l)[-r(l)] \\ \Rightarrow \sum_{l=0}^k \widetilde{P}(l)r(l) &< \sum_{l=0}^k P(l)r(l) \end{aligned}$$

- b. Solution:** For $q \in (0, 1)$, $\widetilde{f}(q) > f(q)$.

Proof:

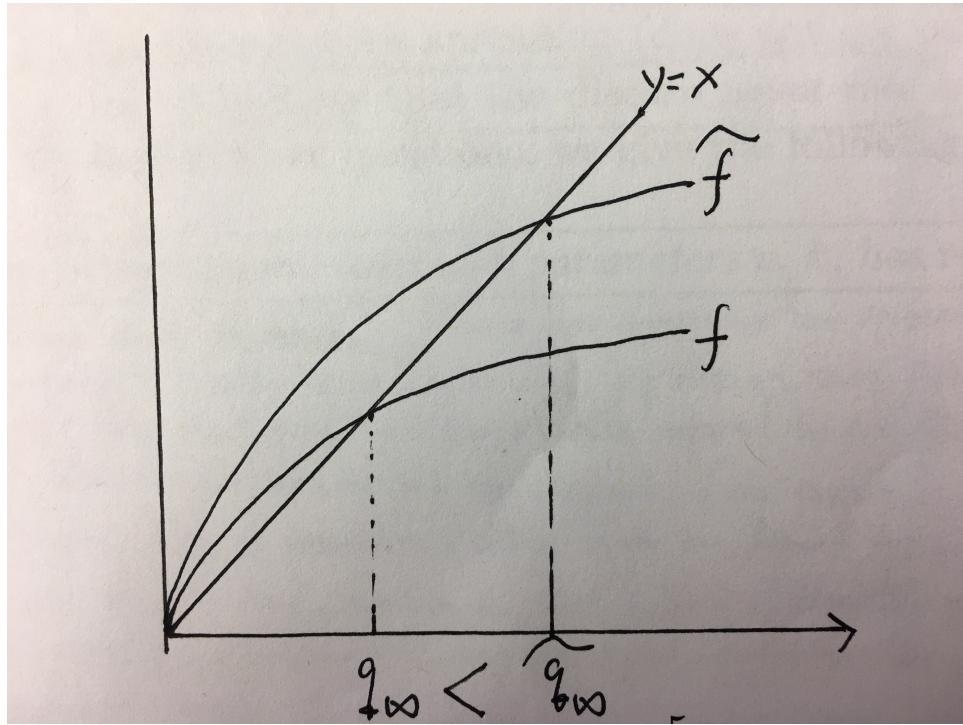
$$f(x) = 1 - \left(1 - \frac{\sum_{i=1}^k iP(i)}{k}x\right)^k$$

$$\widetilde{f}(x) = 1 - \left(1 - \frac{\sum_{i=1}^k i\widetilde{P}(i)}{k}x\right)^k$$

Since \widetilde{P} first-order stochastic dominates P , and the function $r(l) = l$ is strictly increasing, therefore, we have

$$\begin{aligned} \sum_{i=1}^k i\widetilde{P}(i) &> \sum_{i=1}^k iP(i) \\ \Rightarrow \left(1 - \frac{\sum_{i=1}^k i\widetilde{P}(i)}{k}x\right)^k &< \left(1 - \frac{\sum_{i=1}^k iP(i)}{k}x\right)^k \\ \Rightarrow \widetilde{f}(x) &> f(x) \end{aligned}$$

- c. **Solution:** We know from part (b) that, for $x \in (0, 1)$, $\widetilde{f}(x) > f(x)$, that is, the curve $\widetilde{f}(x)$ is always above the curve $f(x)$. When \widetilde{q}_{\inf} and q_{\inf} are both positive, we have the following plot showing the relationship.



Clearly, $\widetilde{q}_{\inf} > q_{\inf}$. It is saying that, the branching process with the first-order stochastically dominant probability distribution \widetilde{P} will never die out with greater positive probability \widetilde{q}_{\inf} than the one with the dominated probability distribution P .

3. (15 points)

- a. **Solution:** Let p_{ij} , q_{ij} , r_{ij} denote the (i,j) entry of the matrices P , Q , and R , respectively. Then,

$$r_{ij} = \sum_{k=1}^n q_{ik} p_{kj}$$

- b. **Solution:** For every $i \in 1, \dots, n$, the sum of all entries in row i of R :

$$\begin{aligned} \sum_{j=1}^n r_{ij} &= \sum_{j=1}^n \left(\sum_{k=1}^n q_{ik} p_{kj} \right) \\ &= \sum_{k=1}^n \left(\sum_{j=1}^n q_{ik} p_{kj} \right) \\ &= \sum_{k=1}^n \left(q_{ik} \sum_{j=1}^n p_{kj} \right) \end{aligned}$$

Since P is a row-stochastic matrix, for every $k \in 1, \dots, n$, $\sum_{j=1}^n p_{kj} = 1$.

$$\begin{aligned} &= \sum_{k=1}^n (q_{ik} \times 1) \\ &= \sum_{k=1}^n (q_{ik}) \end{aligned}$$

Since Q is a row-stochastic matrix, for every $k \in 1, \dots, n$, $\sum_{k=1}^n q_{ik} = 1$.

$$= 1$$

Therefore, the product of two row-stochastic matrices is also a row-stochastic matrix.

- c. **Solution:** Let p_{ij}^k denotes the (i,j) entry of the matrix P^k , where $k \geq 1$. Since P is row-stochastic, from (b) we know that the product of two row-stochastic matrices is also row-stochastic. Therefore, for every $i \in 1, \dots, n$, the sum of all the entries in row i of P^2 is:

$$\sum_{j=1}^n p_{ij}^2 = 1$$

Since P and P^2 are row-stochastic, $P^3 = P \times P^2$ is row-stochastic, therefore, for every $i \in 1, \dots, n$, the sum of all the entries in row i of P^3 is:

$$\sum_{j=1}^n p_{ij}^3 = 1$$

For the same reasoning, for every $k \geq 1$, if P^{k-1} is row-stochastic, then $P^k = P \times P^{k-1}$ is row-stochastic. Therefore, for every $i \in 1, \dots, n$, the sum of all the entries in row i of P^k is:

$$\sum_{j=1}^n p_{ij}^k = 1$$

d. Solution:

$$L = P_1^100 = \begin{bmatrix} 0.232 & 0.185 & 0.195 & 0.245 & 0.144 \\ 0.232 & 0.185 & 0.195 & 0.245 & 0.144 \\ 0.232 & 0.185 & 0.195 & 0.245 & 0.144 \\ 0.232 & 0.185 & 0.195 & 0.245 & 0.144 \\ 0.232 & 0.185 & 0.195 & 0.245 & 0.144 \end{bmatrix}, LP_1 = P_1^101 = \begin{bmatrix} 0.232 & 0.185 & 0.195 & 0.245 & 0.144 \\ 0.232 & 0.185 & 0.195 & 0.245 & 0.144 \\ 0.232 & 0.185 & 0.195 & 0.245 & 0.144 \\ 0.232 & 0.185 & 0.195 & 0.245 & 0.144 \\ 0.232 & 0.185 & 0.195 & 0.245 & 0.144 \end{bmatrix}$$

$$L = P_2^100 = \begin{bmatrix} 0.159 & 0.108 & 0.295 & 0.297 & 0.141 \\ 0.159 & 0.108 & 0.295 & 0.297 & 0.141 \\ 0.159 & 0.108 & 0.295 & 0.297 & 0.141 \\ 0.159 & 0.108 & 0.295 & 0.297 & 0.141 \\ 0.159 & 0.108 & 0.295 & 0.297 & 0.141 \end{bmatrix}, LP_2 = P_2^101 = \begin{bmatrix} 0.159 & 0.108 & 0.295 & 0.297 & 0.141 \\ 0.159 & 0.108 & 0.295 & 0.297 & 0.141 \\ 0.159 & 0.108 & 0.295 & 0.297 & 0.141 \\ 0.159 & 0.108 & 0.295 & 0.297 & 0.141 \\ 0.159 & 0.108 & 0.295 & 0.297 & 0.141 \end{bmatrix}$$

$$L = P_3^100 = \begin{bmatrix} 0.141 & 0.195 & 0.376 & 0.202 & 0.086 \\ 0.141 & 0.195 & 0.376 & 0.202 & 0.086 \\ 0.141 & 0.195 & 0.376 & 0.202 & 0.086 \\ 0.141 & 0.195 & 0.376 & 0.202 & 0.086 \\ 0.141 & 0.195 & 0.376 & 0.202 & 0.086 \end{bmatrix}, LP_3 = P_3^101 = \begin{bmatrix} 0.141 & 0.195 & 0.376 & 0.202 & 0.086 \\ 0.141 & 0.195 & 0.376 & 0.202 & 0.086 \\ 0.141 & 0.195 & 0.376 & 0.202 & 0.086 \\ 0.141 & 0.195 & 0.376 & 0.202 & 0.086 \\ 0.141 & 0.195 & 0.376 & 0.202 & 0.086 \end{bmatrix}$$

I notice that, P_1 , P_2 , P_3 are all row-stochastic, therefore, both $L = P^100$ and $LP = P^101$ are row stochastic, as well. Furthermore, $L = LP$. I think, when n is large enough, for any row-stochastic matrix, there exists a limit, such that, $\lim_{n \rightarrow \infty} P^n = P^*$. In other words, $\exists N_0, s.t., \forall i, j \geq N_0$,

$$P^i = P^j$$

e. Solution: Give a probabilistic interpretation of P^k : View P as the transition matrix of a Markov process on the state space $1, \dots, n$, with p_{ij} being the probability of switching to state j conditional on being in state i , the P^k is the transition matrix of a Markov process after k steps.

Give a probabilistic interpretation of your answer in (d): We have noticed that, if P is a row-stochastic matrix, then there is a long-run steady state p^* of the Markov process, that is, $\lim_{n \rightarrow \infty} P^n = P^*$. In other words, starting from an initial state s_0 , there exists N_0 , such that, after N_0 steps, the system tends to stabilize at the state P^*s_0 , where no further change will happen.

4. Braess' Paradox (Easley and Kleinberg, 8.4 Q1) (15 points)

a. Solution: The Nash equilibrium only occurs when 500 cars take route ACB and 500 cars take route ADB.

Prove that this is a Nash equilibrium: Assume 500 cars take route ACB and 500 cars take route ADB, the payoff for each of these players is

$$u_{x_1}(ACB, s_{-x_1}) = -\left(\frac{x_{ACB}}{100} + 12\right) = -\left(\frac{500}{100} + 12\right) = -17$$

Without loss of generality (since route ACB and ADB are symmetric, players choose either route have the same situation), we can look at player x_1 with strategy ACB. If x_1 switches strategy to ADB, assuming all other players stay the same, player x_1 would have a new payoff

$$u_{x_1}(ADB, s_{-x_1}) = -\left(\frac{x_{ADB}}{100} + 12\right) = -\left(\frac{501}{100} + 12\right) = -17.1 < -17$$

So x_1 would not want to switch strategy. Since we have assumed generality, no player would want to switch strategies. Thus, this is a Nash equilibrium.

Prove that this even split is the only Nash equilibrium: If there were uneven split, any one player in the more congested route could improve his/her payoff by switching to the less congested route, assuming all other players don't change their strategies. Therefore, any uneven split is not a Nash equilibrium. This proves that even split is the only Nash equilibrium.

In this case, the total cost of travel is 17,000.

- b. **Solution:** In the modified scenario, each player now has three options, $S_i = ACB, ADB, ACDB$. The Nash equilibrium occurs when all 1,000 cars take the route ACDB, that is,

$$x = y = 1000$$

Prove that this is a Nash equilibrium: Assume all 1,000 players take route ACDB. Without loss of generality, we can look at the payoff for player x_1 :

$$u_{x_1}(ACDB, s_{-x_1}) = -\left(\frac{x_{AC}}{100} + 0 + \frac{x_{DB}}{100}\right) = -(10 + 0 + 10) = -20$$

If x_1 switches strategy to ACB or ADB, assuming all other players stay the same, then x_1 would have a new payoff of:

$$u_{x_1}(ACB, s_{-x_1}) = -\left(\frac{x_{AC}}{100} + 45\right) = -(10 + 12) = -22 < -20$$

$$u_{x_1}(ADB, s_{-x_1}) = -\left(\frac{x_{AD}}{100} + 45\right) = -(0.01 + 12) = -22.01 < -20$$

So x_1 would not want to switch strategy. Since we have assumed generality, no player would want to switch strategies. Thus, this is a Nash equilibrium.

We can show that an even split is no longer a Nash equilibrium anymore: Assume 500 cars take route ACB and 500 cars take route ADB, for any player x_1 with strategy ACB:

$$u_{x_1}(ACB, s_{-x_1}) = -\left(\frac{x_{ACB}}{100} + 12\right) = -\left(\frac{x_{ADB}}{100} + 12\right) = -17$$

If x_1 switches strategy to ACDB, assuming all other players stay the same, player x_1 would have a new payoff

$$u_{x_1}(ACDB, s_{-x_1}) = -\left(\frac{x_{AC}}{100} + 0 + \frac{x_{DB}}{100}\right) = -\left(\frac{500}{100} + 0 + \frac{501}{100}\right) = -10.01 < -17$$

So x_1 would want to switch strategy. Since ACB and ADB give the same payoff initially, for any players, regardless of what their strategies were, either ACB or ADB, they would improve their payoffs by switching to ACDB, assuming all other players stay the same. This proves that an even split is not a Nash equilibrium.

As a result of the addition of the new road, the total travel of time for the 1,000 cars is 20,000, which is even worse than without the new road CD. Therefore, the result of adding a congestion free road, is a socially suboptimal equilibrium.

c. **Solution:**

The Nash equilibrium only occurs when 500 cars take route ACB and 500 cars take route ADB. In this case, the total cost of travel is 10,000.

Prove that this is a Nash equilibrium: Assume 500 cars take route ACB and 500 cars take route ADB, the payoff for each of these players is

$$u_{x_1}(ACB, s_{-x_1}) = -\left(\frac{x_{ACB}}{100} + 5\right) = -\left(\frac{500}{100} + 5\right) = -10$$

Without loss of generality (since route ACB and ADB are symmetric, players choose either route have the same situation), we can look at player x_1 with strategy ACB. If x_1 switches strategy to ADB, assuming all other players stay the same, player x_1 would have a new payoff

$$u_{x_1}(ADB, s_{-x_1}) = -\left(\frac{x_{ADB}}{100} + 5\right) = -\left(\frac{501}{100} + 5\right) = -10.01 < -10$$

If x_1 switches strategy to ACDB, assuming all other players stay the same, player x_1 would have a new payoff

$$u_{x_1}(ACDB, s_{-x_1}) = -\left(\frac{x_{AC}}{100} + 0 + \frac{x_{DB}}{100}\right) = -\left(\frac{500}{100} + 0 + \frac{501}{100}\right) = -10.01 < -10$$

So x_1 would not want to switch strategy. Since we have assumed generality, no player would want to switch strategies. Thus, this is a Nash equilibrium.

Prove that ACDB is not an NE: Assume that all 1,000 cars take the route ACDB, each player now has three options, $S_i = ACB, ADB, ACDB$. Without loss of generality, we can look at the payoff for player x_1 :

$$u_{x_1}(ACDB, s_{-x_1}) = -\left(\frac{x_{AC}}{100} + 0 + \frac{x_{DB}}{100}\right) = -(10 + 0 + 10) = -20$$

If x_1 switches strategy to ACB or ADB, assuming all other players stay the same, then x_1 would have a new payoff of:

$$u_{x_1}(ACB, s_{-x_1}) = -\left(\frac{x_{AC}}{100} + 5\right) = -(10 + 5) = -17 > -20$$

$$u_{x_1}(ADB, s_{-x_1}) = -(5 + \frac{x_{DB}}{100}) = -(5 + 10) = -17 > -20$$

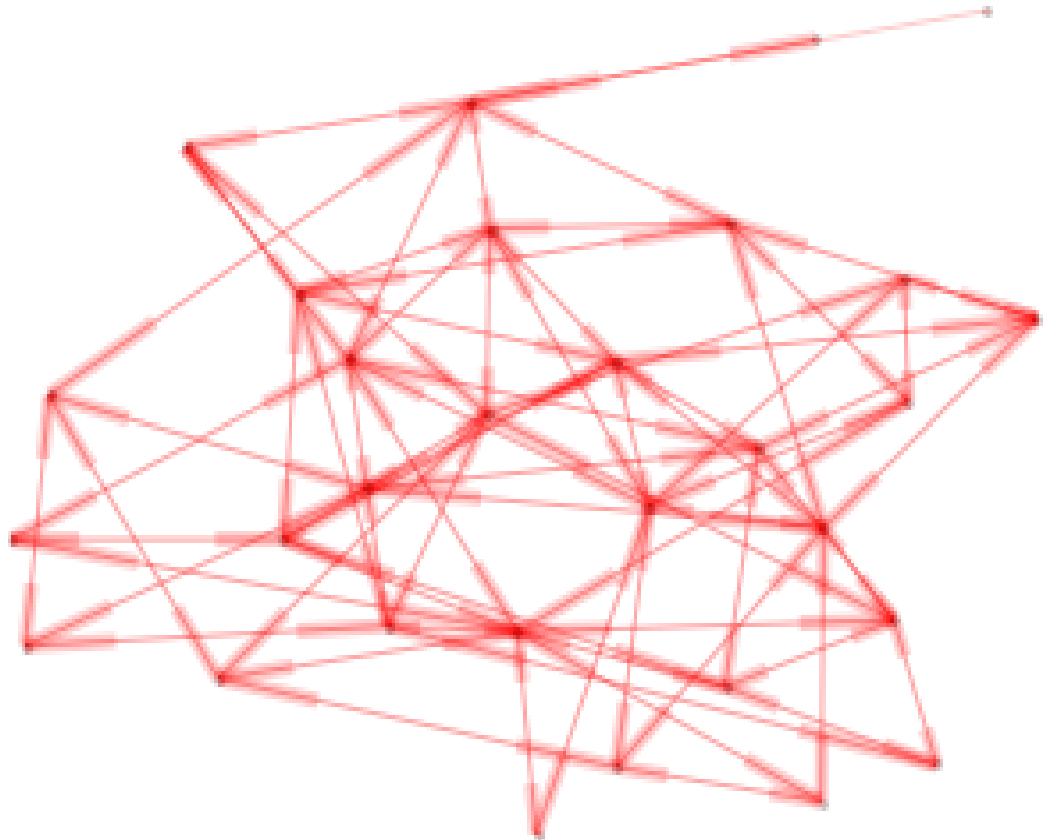
So x_1 would want to switch strategy to either ACB or ADB. Thus, this is not a Nash equilibrium. Furthermore, any strategies involve only taking routes ACB and ADB but not even split, is not a NE, since players taking more congested route can always improve payoff by switching to less congested route, given all other players stay the same.

This proves that even split is the only Nash equilibrium.

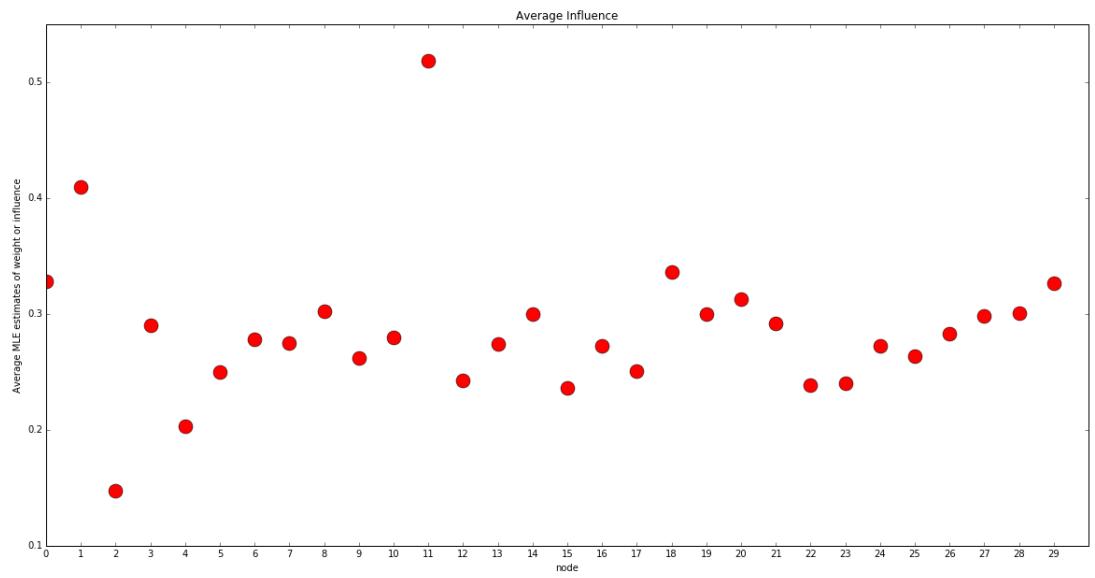
If the government closes the road from C to D, the NE and total cost of travel won't change, since no player takes the road C to D at the initial equilibrium when CD is still there. The total cost of travel is still 10,000.

5. Learning Influence Locally (20 points)

- a. **Solution:** $\tau = 31$. There are 100 cascades. At timestep 2 (using 0-indexing), node 13 (the 14th node) first have its opinion activated in cascade 6 (the 7th cascade).
- b. **Solution:** There are 30 nodes in the network, and average out-degree is 5.23. A visualization of the network as follows:



- c. **Solution:** For argument $(10,4,1,opinion_0)$, my funciton output is 1.
- d. **Solution:** The estimated weight of edge $(1,2)$ is 0.67; the estimated weight for edge $(26,21)$ is 0.30. Please see the complete estimates at "MLE.csv".
- e. **Solution:** Node 11 has the highest average edge weight for outgoing edges (0.5185), and node 2 has the lowest average edge weight for outgoing edges(0.1477).



(Only for those not activated initially) On average, node 21 is activated first (average activated time is 1.140845), and node 2 is activated last (average activated time is 1.880597). But node 11 was activated initially for all cascades.

